

Міністерство освіти і науки України

Національний університет водного господарства та  
природокористування

Кафедра екології, технології захисту навколишнього середовища та  
лісового господарства

**05-02-333М**

### **МЕТОДИЧНІ ВКАЗІВКИ**

до виконання практичних робіт з навчальної дисципліни  
«Екологічна статистика» для здобувачів вищої освіти першого  
(бакалаврського) рівня та другого (магістерського) рівня усіх  
освітньо-професійних програм спеціальностей НУВГП  
денної і заочної форм навчання

Схвалено науково-методичною  
радою НУВГП  
Протокол № 4  
від 23.06.2021 року

Рівне – 2021

Методичні вказівки до виконання практичних робіт з навчальної дисципліни «Екологічна статистика» для здобувачів вищої освіти першого (бакалаврського) рівня та другого (магістерського) рівня усіх освітньо-професійних програм спеціальностей НУВГП денної і заочної форм навчання [Електронне видання] / Буднік З. М. – Рівне : НУВГП, 2021. – 36 с.

Укладач: Буднік З. М., к.с.-г.н., доцент кафедри екології, технології захисту навколишнього середовища та лісового господарства.

Відповідальний за випуск: Клименко М. О., доктор с.-г. наук, професор, завідувач кафедри екології, технології захисту навколишнього середовища та лісового господарства.

Вчений секретар  
науково-методичної ради

Костюкова Т. А.

© Буднік З. М., 2021  
© НУВГП, 2021

## ЗМІСТ

### Практична робота №1

статистичні функції. Ранжування даних ..... 4

### практична робота №2

Статистичні методи вивчення динаміки і прогнозування процесів.

Прогнозування методом ковзних середніх значень показників.

Прогнозування методом експоненціального згладжування ..... 7

### Практична робота №3

Трендові моделі ..... 12

### практична робота №4

Методика регресійного аналізу ..... 17

### практична робота №5

Графічне подання даних ..... 27

### Практична робота №6

Використання методів статистичного оброблення даних ..... 30

Список використаної літератури ..... 36

## ПРАКТИЧНА РОБОТА №1

### СТАТИСТИЧНІ ФУНКЦІЇ. РАНЖУВАННЯ ДАНИХ

**Мета:** ознайомитись із використанням статистичних функцій та навчитися використовувати ранжування отриманих статистичних даних

#### Основні поняття

Ранжування великих масивів числових статистичних даних забезпечує їх упорядкування і отримання такої статистичної оцінки, як ранг.

**Ранг** — номер значення, що присвоюється випадковій величині в ряду даних.

Для цього є такі правила:

- кожне унікальне значення має унікальний ранг;
- група чисел з однаковими значеннями ( $m$ ) одержує однаковий ранг, що дорівнює рангу першого числа групи ( $R$ ). Число, що йде за цією групою, одержує ранг, який дорівнює ( $R + m$ );
- максимальне значення ряду має ранг 1, мінімальне значення — ранг, що дорівнює кількості чисел у ряду даних або числу ( $n-m+1$ ), якщо мінімальне значення одержують  $m$  чисел ряду (де  $n$  — загальна кількість чисел у ряду даних).

Значення рангу дає змогу знайти відсоток чисел у множині, які не перевищують заданий ранг, — відсоток-ранг. Останній визначається

для різних чисел за формулою  $\frac{n-R}{n-1} \times 100\%$ , а для однакових чисел

— за формулою  $\frac{n-R-m+1}{n-1} \times 100\%$

Окреме значення показника оцінюється його відносним положенням у ряду даних за допомогою таких статистичних показників:

- квартиля — значення ознаки, що ділить ряд на 4;
- квінтеля — значення ознаки, яке ділить ряд на 5;
- дециля — значення ознаки, що ділить ряд на 10;
- персентиля — значення ознаки, яке ділить ряд на 100.

#### Ранжування числових рядів даних

Для ранжування числових рядів можна використовувати інструмент Ранг и персентиль із «Пакета аналізу».

Вимоги до створення діапазону вхідних даних такі самі, як і для інструменту «Описова статистика».

Діалогове вікно інструменту **Ранг и персентиль** передбачає завдання вхідного діапазону даних, групування, міток і вибір параметрів для виведення результативної таблиці.

Результативна таблиця ранжування включає чотири стовпці для кожного вхідного ряду статистичних показників:

- точка — місце даних (порядковий номер) у вхідній таблиці;
- рядок1\стовпець1 — значення чисел у ряду даних;
- ранг — ранг числа;
- відсоток — відсоток даних у масиві, що не перевищують задане число (відсоток-ранг).

Для ранжування статистичних рядів даних можна також використовувати статистичні функції.

**РАНГ** — обчислює ранг числа в масиві чисел. Синтаксис: РАНГ(число;посилання;порядок):

- число — це число, для якого визначається ранг;
- посилання — це масив або посилання на ряд даних. Нечислові значення в посиланні ігноруються;
- порядок — це число, що визначає метод упорядкування. Якщо порядок дорівнює 0 або опущений, то Microsoft Excel визначає ранг числа так, ніби посилання було списком, відсортованим у порядку спадання. Якщо порядок — це будь-яке ненульове число, то Microsoft Excel визначає ранг числа так, ніби посилання було списком» відсортованим у порядку зростання.

Зауваження. РАНГ присвоює однаковим числам однаковий ранг. Проте наявність однакових чисел впливає на ранг наступних чисел. Наприклад, для списку даних, якщо число 10 з'являється двічі й має ранг 5, то число 9 матиме ранг 7, а жодне число не матиме ранг 6. Наприклад, якщо клітинки A1:A7 містять числа 7, 3, 4, 3, 5, 1, 2 відповідно, то

РАНГ(A1; A\$1: A\$7;1) - 7; РАНГ(A2; A\$1: A\$7;1) = 3.

Знак \$ дає змогу зафіксувати посилання на діапазон клітинок A1:A7 (фіксуються від змін рядки) і здійснювати копіювання формули (наприклад, з клітинки B1 у діапазон клітинок B2:B7).

ПРОЦЕНТРАНГ — визначає питому вагу даних, що не перевищують за значенням певне число ряду даних. Ця функція слугує для оцінювання відносного положення точки у множині даних. Наприклад, можна використовувати функцію ПРОЦЕНТРАНГ, щоб оцінити положення певного результату серед усіх результатів тестування.

Синтаксис: ПРОЦЕНТРАНГ(масив; х; розрядність):

- масив — це масив або інтервал даних із числовими значеннями;
- х — точка даних, для якої визначається відсоток-ранг;
- розрядність — необов'язкове значення, що визначає кількість значущих цифр значення відсоток-рангу. Якщо цей аргумент опущений, то функція ПРОЦЕНТРАНГ має три цифри після десяткової коми.

Наприклад,

$\text{ПРОЦЕНТРАНГ}(\{1;2;3;4;5;6;7;8;9;10\};4) = 0,333$ .

ПЕРСЕНТИЛЬ — обчислює й-й персентиль для значень з інтервалу. Ця функція слугує для визначення порогу прийнятності. Наприклад, можна прийняти рішення екзамінувати тільки тих кандидатів, які набрали балів більш як 90-й персентиль.

Синтаксис: ПЕРСЕНТИЛЬ(масив;к):

- масив — це масив або інтервал даних із числовими значеннями;
- к — значення персентилів в інтервалі від 0. до 1 включно.

Наприклад,

$\text{ПЕРСЕНТИЛЬ}(\{1;2;3;4\};0,3) = 1,9$ .

КВАРТИЛЬ — обчислює квартиль множини даних. Квартиль часто використовують при аналізі продажу, щоб розбити генеральну сукупність даних на групи. Наприклад, можна скористатися функцією КВАРТИЛЬ, щоб знайти 25 % найбільш дохідних підприємств серед усієї їх множини.

Синтаксис: КВАРТИЛЬ(масив;частина):

- масив – це масив або інтервал клітинок із числовими значеннями, для визначається значення квартилів;
- частина – значення яке потрібно обчислити, причому:

0 – мінімальне значення;

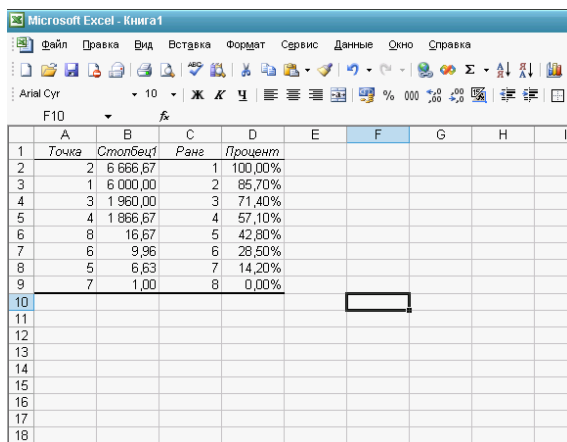
1 – перший квартиль (25-й персентиль);

2 – значення медіани (50-й персентиль);

3 – третій квартиль (75-й персентиль);

4 – максимальне значення

Наприклад, КВАРТАЛЬ({1;2;4;7;8;9;10;12};1)=3,5



	A	B	C	D	E	F	G	H	I
1	Точка	Столбець	Рядок	Процент					
2		2	6 666,67	1	100,00%				
3		1	6 000,00	2	85,70%				
4		3	1 960,00	3	71,40%				
5		4	1 866,67	4	57,10%				
6		8	16,67	5	42,80%				
7		6	9,96	6	28,50%				
8		5	6,63	7	14,20%				
9		7	1,00	8	0,00%				
10									
11									
12									
13									
14									
15									
16									
17									
18									

## ПРАКТИЧНА РОБОТА №2

### СТАТИСТИЧНІ МЕТОДИ ВИВЧЕННЯ ДИНАМІКИ І ПРОГНОЗУВАННЯ ПРОЦЕСІВ. ПРОГНОЗУВАННЯ МЕТОДОМ КОВЗНИХ СЕРЕДНІХ ЗНАЧЕНЬ ПОКАЗНИКІВ. ПРОГНОЗУВАННЯ МЕТОДОМ ЕКСПОНЕНЦІАЛЬНОГО ЗГЛАДЖУВАННЯ

**Мета:** навчитися здійснювати прогнозування змін екологічного стану навколишнього середовища з використанням статистичних показників

#### Основні поняття

Ряд розміщених у хронологічній послідовності значень статистичних показників називається часовим або динамічним.

Кожний такий ряд включає два елементи:

- моменти або періоди часу, яких стосуються статистичні дані, що наводяться;
- статистичні показники, які характеризують об'єкт чи явище за певний момент часу або за вказаний період.

Динамічні ряди характеризують процеси розвитку соціально-економічних явищ. Цим процесам властиві дві взаємопов'язані риси: динамічність та інерційність. Перша проявляється зміною значень і

варіацією показників, що характеризують процес; друга — сталістю механізму формування процесу, напрямом та інтенсивністю розвитку протягом певного часу. Від поєднання цих рис динамічний ряд у будь-який момент міститиме залишки минулого, основи сучасного і зародки майбутнього.

Діалектична єдність мінливості та сталості, динамічності й інерційності формує закономірності розвитку процесу. Під впливом різних чинників довго- і короткострокової дій в одних рядах значення показників протягом тривалого часу зростають або спадають з різною інтенсивністю, а в інших збільшення та зменшення значень показників чергуються з певною періодичністю. Окрім закономірних коливань значень показників, динамічним рядам також можуть бути притаманні випадкові коливання.

При моделюванні динамічних процесів причинний механізм формування властивих їм особливостей в явному вигляді не враховується. Будь-який процес розглядається як функція часу. Хоча час не є чинником конкретного соціально-економічного процесу, проте зміна показників у часі, як правило, акумулює комплекс постійно діючих умов і причин, що визначають цей процес. Виявлення основної тенденції розвитку процесу називається вирівнюванням **динамічного ряду**, а методи її виявлення — методами вирівнювання. Останнє дає змогу характеризувати особливості змін у часі динамічного ряду в найзагальнішому вигляді як функцію часу. При цьому вважається, що через час можна виразити вплив усіх основних чинників.

### ***Прогнозування методом ковзних середніх значень показників***

Досить поширеним і простим методом аналізу динамічного ряду є його згладжування. При цьому фактичні значення показників замінюються їх середніми значеннями за певний інтервал. Варіація середніх значень показників порівняно з варіацією емпіричного ряду значно менша, а тому характер динаміки процесу проявляється чіткіше. Лінійне згладжування динамічного ряду можна здійснювати методом ковзних середніх значень показників.

Цей метод застосовується для прогнозування процесів з незначною варіацією середніх значень показників протягом коротких періодів часу. Всі спостереження часового ряду мають



однакову значущість  $1/n$  незалежно від їхнього місця у вхідних даних. При досить великих  $n$  ( $n > 25$ ) додання нових даних майже не змінює їх середнє значення для попереднього моменту часу.

Для визначення ковзного середнього значення показників формують збільшені інтервали, що складаються з однакової кількості періодів. Кожний наступний інтервал одержують, послідовно просуваючись від початкового значення динамічного ряду на один крок уперед. Ковзне середнє значення показників використовують для розрахунку їхніх значень у прогнозованому періоді на основі середнього значення показників для зазначеного числа попередніх періодів. Кожне прогнозоване значення

показників визначається виразом 
$$\bar{y}_{(t)} = \frac{\sum_{i=t-m+1}^t y_{(i)}}{m},$$

де  $t$  — число попередніх періодів;  $y_{(i)}$  — фактичне значення показників у момент часу  $i$ ;  $y_t$  — прогнозоване значення показників у момент часу  $t$ .

Отже, перший інтервал включатиме значення  $y_1, y_2, \dots, y_m$ ; другий — значення  $y_2, y_3, \dots, y_{m+1}$  і т. д. Таким чином, інтервал згладжування ніби ковзає по динамічному ряду з кроком, що дорівнює одиниці. Одержані середні значення показників належать до середини інтервалу, тому технічно зручніше формувати збільшений інтервал з непарної кількості періодів (3, 5 тощо). При парному  $t$  середина інтервалу лежатиме між двома часовими точками, і тоді виконують додаткову процедуру центрування (усереднення) кожної пари значень.

Ковзне середнє значення, на відміну від простого середнього для всієї вибірки, містить інформацію про тенденцію зміни даних. Процедура може використовуватися для прогнозування збуту, випуску продукції та інших процесів.

Найефективніше цей метод дослідження можна провести за допомогою інструменту **Скользящее среднее** з «Пакета аналізу».

Параметри діалогового вікна цього інструменту передбачають:

- задання вхідного інтервалу даних динамічного ряду (можна одночасно задавати один рядок або стовпець, тобто один показник, що включає не менш як чотири значення);

- введення міток у першому рядку (стовпці), якщо у вхідний діапазон увійшла клітинка з назвою показника (коли назву виділено, а прапорець міток не активізовано, виникне помилка в

нечислових даних; незалежно від уведення міток Excel їх ігнорує, і потрібно додати назву показника вручну);

- задання інтервалу для визначення ковзного середнього значення (за замовчуванням інтервал дорівнює 3);

- задання вихідного діапазону — верхньої лівої клітинки діапазону для виведення результату. (Оскільки користувач створюватиме стовпець чи рядок назв показників самостійно, більш правильним буде задання не першої, а другої лівої клітинки, щоб у крайній лівій (верхній) клітинці можна було б створити стовпець (рядок) назв показників. Вихідний діапазон та вхідні дані мають знаходитися на одному листку);

- задання виведення автоматично створеного графіка, що відображає вхідні й розрахункові значення динамічного ряду даних;

- виведення додаткового рядка (стовпця) стандартних похибок.

Результати визначення тенденції розвитку виробничих процесів та реалізації продукції показано на рис. 2.

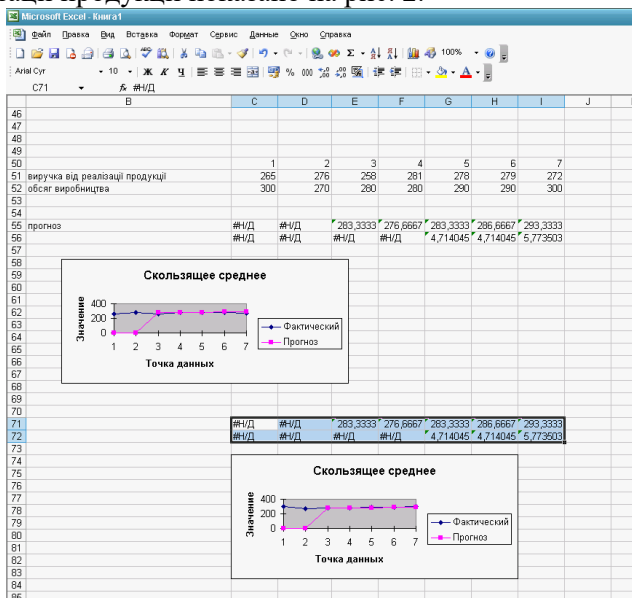


Рис. 2. Порівняння прогнозованих статистичних показників, визначених за методом ковзних середніх значень, із фактичними даними

Видно зростання обох показників і відсутність їх випадкових коливань (графіки фактичних та прогнозованих значень показників майже накладаються). Водночас динамічний ряд обсягу виробництва має чіткішу тенденцію зростання, про що свідчить також менша похибка ряду.

Дослідження основної тенденції розвитку процесів методом ковзних середніх значень показників є емпіричним способом їх попереднього аналізу. Цей метод виступає як допоміжний для спрощення використання інших методів. Його основна перевага — наочність і простота визначення тенденції. Крім того, ковзні середні значення показників залежать від даних за попередні періоди, а при значних інтервалах згладжування може бути виявлена циклічність процесу навіть за відсутності циклів у динамічному ряду (ефект Слущкого).

#### ***Прогнозування методом експоненціального згладжування***

Експоненціальні середні значення показників мають більшу часову стійкість порівняно з ковзними середніми значеннями. В їх визначенні враховуються всі спостереження вхідного ряду даних, однак з різними ваговими коефіцієнтами. При цьому залежність чергових прогнозованих значень показників від попередніх більш сильна, ніж від значень у початковий період. Емпіричний часовий ряд даних визначається виразом

$$y'_t = (1 - \alpha)y'_{t-1} + \alpha y_t$$

де  $y'_t$  — прогнозоване значення показників для періоду  $t$ ;  $y'_{t-1}$  — те саме, але для періоду  $(t-1)$ ;  $y_t$  — фактичне значення показників для періоду  $t$ ;  $0 < \alpha < 1$  — коефіцієнт експоненціального згладжування, що визначає вагу  $t$ -го періоду.

Чим більше значення  $\alpha$ , тим сильніший вплив коливань ряду даних, і навпаки. При невеликих значеннях  $\alpha$  процес є більш інерційним до фактичних значень показників і більше враховуються їх прогнозовані значення за попередній період.

«Пакет аналізу» Excel включає спеціальний інструмент Експоненциальное сглаживание. Виконання цієї процедури аналогічне процедурі Скользящее среднее.

На відміну від деяких інших інструментів «Пакета аналізу», результати (таблиці та графіки), одержувані за допомогою інструментів Експоненциальное сглаживание і Скользящее среднее,

є динамічними (зберігають зв'язки з вхідними даними й автоматично змінюються при коригуванні вхідних даних динамічних рядів). Хоч одержати ці результати можна лише на тому самому робочому листку, що й вхідні дані, проте користувач може відформатувати і перемістити їх за необхідності на інший робочий листок або в іншу робочу книгу.

	A	B	C	D	E	F	G	H	I	J	K	L											
13		Середньорічний виробіток на одного працюючого	грн./чол.	6000	5192,31	5833,38	5490,2	5576,92	5576,92	6000													
16		Середньорічний виробіток на одного робітника	грн./чол.	6666,67	6136,36	6511,63	6086,96	6304,35	6444,44	6666,67													
17		Середньорічна заробітна плата одного працюючого	грн./чол.	1960	1961,64	1958,33	1960,78	1884,62	1942,31	1920													
18		Середньорічна заробітна плата одного робітника	грн./чол.	1866,67	1931,82	1906,98	1913,04	1869,57	1911,11	1866,67													
19		Прибуток від реалізації продукції	тис. грн.	24	24	1	13	27	18	24													
20		Рентабельність підприємства	%	6,63	6,03	0,25	3,18	6,35	4,35	5,97													
21		Рентабельність витрат	%	9,96	9,52	0,39	4,85	10,76	6,9	9,68													
22		Фондовіддача основних фондів	грн./грн.	1	0,82	0,85	0,82	0,85	0,84	0,91													
23		Фондовіддача нематеріальних активів	грн./грн.	16,67	12,86	14	12,73	13,18	12,61	13,64													
24																							
25																							
26	#И/Д	270	260	240	190	90																	
27	#И/Д		#И/Д	#И/Д	31,62278	65,57439																	
28																							
29																							
30																							
31																							
32		<div>Экспоненциальное сглаживание</div> <div><div><div>Значение</div><div>400</div><div>200</div><div>0</div></div><div><div>1</div><div>2</div><div>3</div><div>4</div><div>5</div><div>6</div></div><div><div>Точка данных</div></div></div> <div><div>● Фактический</div><div>● Прогноз</div></div>																					
33																							
34																							
35																							
36																							
37																							
38																							
39																							

## ПРАКТИЧНА РОБОТА №3 ТРЕНДОВІ МОДЕЛІ

**Мета:** навчитись застосовувати трендові моделі при оцінці стану навколишнього середовища

### Основні поняття

Значення деяких економічних показників, що мають випадковий характер, можна використовувати для побудови часових рядів — емпіричної послідовності даних, здобутих у певні моменти часу. Кожний такий ряд характеризується деякою тенденцією розвитку процесу в часі, яка називається трендом. Трендові моделі часових (динамічних) рядів

забезпечують видачу прогнозів на коротко- та середньостроковий періоди при виконанні низки умов:

- період часу, за який досліджується прогнозований процес, має бути достатнім для виявлення закономірностей;
- трендова модель в аналізований період має розвиватись еволюційно;
- процес, що описується часовим рядом даних, повинен мати певну інерційність, тобто для великих змін у поведінці процесу потрібний значний час;
- автокореляційна функція часового ряду даних і його залишкового ряду мусить бути швидко згасаючою, тобто вплив більш пізньої інформації має сильніше відображатись на прогнозованій оцінці, ніж вплив більш ранньої інформації.

На практиці найпоширенішими методами статистичного дослідження тренда є:

- збільшення інтервалів для визначення тренда в часових рядах даних, що коливаються;
- метод ковзних середніх значень із заданим періодом  $m$ ;
- метод аналітичного вирівнювання у вигляді функції тренда, яка залежить від часу.

Лінія тренда широко застосовується для розв'язання задач прогнозування за допомогою методів регресійного аналізу. Підбір функції тренда здійснюється методом найменших квадратів. Для оцінювання точності моделі використовують коефіцієнт детермінації, побудований на основі оцінок дисперсії емпіричних даних та значень трендової моделі.

Трендова модель показує тенденцію розвитку процесу, якщо коефіцієнт детермінації прямує до 1.

Явища, що спостерігаються у часі, можуть розвиватися так:

- рівномірно при сталому абсолютному прирості чергового рівня часового ряду даних за лінійним законом

$$y = a_0 + a_1 t,$$

де  $a_0$  — стала;  $a_1$  — коефіцієнт регресії, що визначає швидкість і напрямок ( $< 0$  — спадання;  $> 0$  — зростання) розвитку;

- рівноприскорено при сталому в часі збільшенні (зниженні) темпу приросту рівнів за законом (парабола другого порядку)

$$y = a_0 + a_1 t + a_2 t^2,$$

де  $a_2$  — коефіцієнт, що характеризує сталу зміну швидкості (темпу) розвитку ( $a_2 > 0$  — прискорення розвитку,  $a_2 < 0$  — його сповільнення);

- із змінним прискоренням (сповільненням) при змінному в часі збільшенні (зменшенні) розвитку за законом (парабола третього — шостого порядків)

$$y = a_0 + a_1t + a_2t^2 + a_3t^3 \dots + a_6t^6;$$

- зі сповільненням зростання в кінці періоду, коли приріст у кінцевих значеннях ряду даних прямує до нуля за законом (логарифмічна функція)

$$y = a_1 \ln t + a_0;$$

- зі зростанням за експоненціальним законом

$$y = a_0 e^{a_1 t},$$

де  $t$  — випадковий час появи чергової події;

- зі сталим відносним приростом за законом степеневі функції (гіпербола)

$$y = a_0 t^{a_1}, y = a_0 + a_1/t.$$

Microsoft Excel будує трендові моделі графічним способом на основі двовимірних діаграм: лінійних, графіків, гістограм, точкових, що відображають динамічні зміни.

Послідовність виконання цієї процедури така:

- будують діаграму;
- діаграму переводять у режим редагування;
- виділяють ряд на діаграмі для побудови лінії тренда;
- подають команду **Додати лінію тренда...** з меню

**Діаграма** або за допомогою контекстного меню.

У результаті на екрані монітора з'являється діалогове вікно, у першому розділі якого можна визначити тип лінії тренда (лінійний, логарифмічний, поліномний, степеневий, експоненціальний, ковзних середніх значень), а у другому задати її параметри:

- ім'я (автоматично з назвою трендової моделі або ввести у текстове поле);

- кількість періодів прогнозування наперед (проводиться на 0,5; 1; 1,5 і т. д. періоди, точний прогноз може здійснюватись тільки на невеликий період, особливо якщо масив фактичних даних невеликий);

- кількість періодів прогнозування назад;

- У-перетин — точку, в якій лінія тренда має перетинати вісь У;

•  $R^2$  — виведення коефіцієнта детермінації, а також відобразити рівняння лінії тренда на діаграмі.

Спочатку, як правило, вибирають трендову модель і задають її параметри (рівняння,  $R^2$ ). У подальшому підбирають тип тренда ( $R^2 \rightarrow 1$ ), двічі клацнувши лівою клавішею миші по лінії тренда, та перевіряють іншу трендову модель (рис. 3).

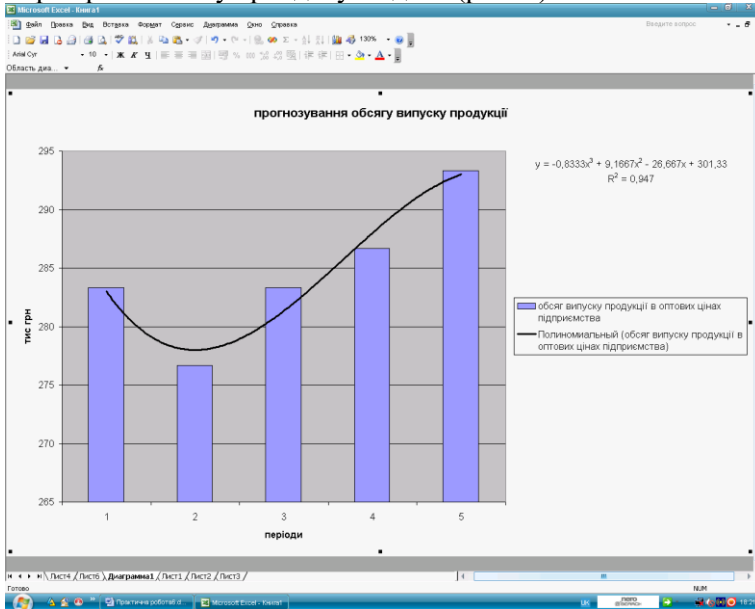


Рис. 3. Прогнозування обсягу випуску продукції за допомогою лінії тренда

Стосовно даних рис. 3 найкращою є модель полінома четвертого степеня. На діаграмі можна спостерігати нерівну тенденцію зростання обсягу випуску продукції у восьмому періоді. Щоб одержати прогнозоване значення обсягу випуску, можна підставити часовий період (8) у рівняння тренда. В результаті дістанемо приблизно 844 тис. грн.

### ***Прогнозування за допомогою статистичних функцій***

Ексел надає широкі можливості здійснювати прогнозування за допомогою статистичних функцій. Розглянемо процедуру їх використання.

**ТЕНДЕНЦИЯ** – обчислює прогнозоване значення відповідно до лінійного тренда; апроксимує прямою лінією (за методом

найменших квадратів) масив Известные значения у та Известные значения х. обчислює значення відповідно до цієї прямої для заданого масиву Новые значения х.

#### Хід роботи

1. Використовуючи інструмент Описательная статистика визначити статистичні показники за вихідними даними (результати вивести на листок 2). Оформити результати у вигляді таблиці, вилучивши дублюючі елементи.

Додаткові параметри:

- рівень надійності – 85%;
  - К-й найменший – 2;
  - К-й найбільший – 2;
  - Підсумкова статистика - задіяна.
2. використовуючи статистичні функції, визначити (на листку 3) такі статистичні показники за вхідними даними:
    - середнє арифметичне;
    - стандартне відхилення;
    - дисперсію;
    - мінімум;
    - максимум;
    - ранг.
  3. Визначити ранг за вхідними показниками і вивести результати на листок 4, користуючись інструментом Ранг «Пакета аналізу».
  4. Спрогнозувати методом ковзних середніх перші два статистичні показники до другого завдання. Результати вивести на листок 4. параметри:
    - інтервал – 2;
    - виведення графіка та стандартних помилок – задіяно.
  5. Спрогнозувати методом експоненціального згладжування третій і четвертий показники до другого завдання. Результати вивести на листок 5.

Параметри:

- чинник згасання – 0,6;
  - виведення графіка та стандартних помилок - задіяно.
6. Побудувати на листку 6 різні діаграми динамічної зміни вхідних показників і вивести лінію тренда. Підібрати тип цієї



лінії, та спрогнозувати зміну показників на два кроки вперед. На діаграму занести рівняння лінії тренда, коефіцієнт детермінації, а також плани помилок.

## **ПРАКТИЧНА РОБОТА №4**

### **МЕТОДИКА РЕГРЕСІЙНОГО АНАЛІЗУ**

**Мета:** навчитись використовувати методику регресійного аналізу при оцінці показників екологічної підсистеми

#### **Основні поняття**

При визначенні конкретних залежностей одні показники розглядаються як фактори впливу (ознаки), Що обумовлюють зміни іншого показника (результативного фактора). Функціональні зв'язки характеризуються повною відповідністю між змінами факторної ознаки і змінами результативної величини, причому кожному значенню фактора-ознаки відповідає певне значення результативного фактора. При кореляційних зв'язках між змінами факторів-ознак та результативного показника повної відповідності не існує. Вплив окремих факторів виявляється лише в середньому при значній кількості спостережень фактичних даних. Крім того, фактор-ознака, як правило, залежить від зміни інших показників.

Форма взаємозв'язку випадкових величин і функції дістала назву рівняння регресії. Виділяють парну (просту) та множинну регресії лінійного і нелінійного (квадратичного, експоненціального, напівлогарифмічного) типів. Вид, а також параметри рівняння регресії знаходять за допомогою Методу найменших квадратів. За наявності кореляційної залежності визначають лише тенденцію зміни результативного показника при змінах факторів-ознак.

Найчастіше застосовуються такі математичні залежності для оцінювання кореляційного зв'язку між факторами:

- прямолінійна
- параболічна
- показникова
- степенева
- гіперболічна
- напівлогарифмічна

Статистичне оцінювання тісноти зв'язку ґрунтується на показниках варіації:

- загальній дисперсії  $\sigma_y^2$  результативного показника,

обумовлений впливом усіх факторів у сукупності;

- факторній дисперсії  $\sigma_y^2$  результативного показника, що показує його варіацію під впливом окремих факторів;

залишковій дисперсії  $\sigma_y^2$  результативного показника, яка показує його варіацію під впливом усіх факторів, крім виділеного.

Якісною оцінкою ступеня зв'язку випадкових величин виступає коефіцієнт детермінації, що визначається виразом  $R^2 = \frac{\sigma_{yx}^2}{\sigma_y^2}$  —

відношенням факторної та загальної дисперсій. Індекс кореляції розраховується як квадратний корінь із коефіцієнта детермінації, тобто  $R = \sqrt{R^2}$ , причому його значення лежать у межах від -1 до +1 (знак «мінус» указує на наявність зворотного зв'язку між факторами).

Для оцінювання значущості індексу кореляції можна використовувати ***F - критерій Фішера***.

$$F_R = \frac{R^2}{1 - R^2} \frac{n - m}{m - 1},$$

де  $n$  — число значень у масиві;  $m$  — число параметрів рівняння регресії (факторів).

Фактичне значення цього критерію порівнюють із критичним значенням, яке визначають з урахуванням рівня значущості та кількості ступенів вільності. Якщо фактичне значення  $F$ -критерію Фішера більше від критичного, то індекс кореляції  $R$  вважається істотним.

Якісне оцінювання ступеня зв'язку випадкових величин може бути виконане з використанням коефіцієнта детермінації за шкалою Чеддона:

0,1...0,3 — незначний; 0,7 ... 0,9 — високий;

0,3...0,5 — помірний; 0,9 ... 0,99 — дуже високий;

0,5...0,7 — істотний; 1,0 — функціональний.

Якщо коефіцієнт детермінації  $R^2 > 0,7$ , то варіація залежної змінної в основному обумовлена впливом факторів, і для прогнозування можна використовувати одержані регресійні моделі.

Якщо аналізується невелика сукупність даних ( $n < 30$ ), то для визначення їх довірчого інтервалу використовується ***t-критерій***

**Стьюдента.** Розраховане значення  $t$  – критерію  $t_p = \frac{R\sqrt{n-2}}{1-R^2}$  для

коефіцієнта кореляції порівнюється з критичним з урахуванням прийнятого рівня значущості, а також кількості ступенів вільності та вважається типовим, якщо  $t_p > t_k$ . Аналогічно оцінюється значущість факторів  $x$  на основі  $t$ -критерію Стьюдента залежно від кількості ступенів вільності  $k=n-2$  і рівня значущості наведено в табл. 1.

Табличний редактор Excel дає змогу використати різні інструменти «Пакета аналізу» (**Корреляция, Ковариация, Регрессия**) для одержання параметрів лінійної парної та множинної регресій, а також оцінки ступеня зв'язку.

**Зауваження.** Під час проведення кореляційно-регресійного аналізу слід ураховувати такі вимоги до вхідних даних, щоб одержати вірогідні результати:

- статистична сукупність даних має включати достатню кількість спостережень або однорідних об'єктів (не менш як п'ять) — чим більша кількість спостережень, тим точнішими будуть результати одержаних рівнянь залежності;
- статистичні дані мають бути відібрані за однакові періоди часу (місяць, квартал, рік) або для однорідних об'єктів;
- при проведенні множинної регресії кількість факторів має бути меншою (хоча б на два), ніж кількість спостережень.

$\infty$	1,645	1,960	2,241	2,326	2,576	2,807	2,968	3,090	3,291
----------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Розглянемо процедуру проведення багатофакторного кореляційного аналізу із застосуванням інструментів «Пакету аналізу» Excel.

**Етап 1.** Визначаємо фактори, що впливають на результативний показник, і відбираємо найістотніші з них. Основні правила відбору факторів:

- результативним фактором, як правило, визначається якісний показник ефективності певної сфери діяльності підприємства (прибуток, рентабельність, дохідність, обсяг реалізації, собівартість тощо);

Таблица 1

### Значення t-критерію Стюдента

$\alpha \backslash k$	0,9	0,95	0,975	0,98	0,99	0,995	0,997	0,998	0,999
1	2	3	4	5	6	7	8	9	10
1	6,13	12,71	25,45	31,82	63,66	127,3	212,2	318,3	636,6
2	2,920	4,303	6,205	6,965	9,925	14,09	18,27	22,33	31,60
3	2,353	3,182	4,177	4,541	5,841	7,453	8,891	10,21	12,92
4	2,132	2,776	3,495	3,747	4,604	5,597	6,435	7,173	8,610
5	2,015	2,571	3,163	3,365	4,032	4,773	5,376	5,893	6,869
6	1,943	2,447	2,969	3,143	3,707	4,317	4,800	5,208	5,959
7	1,696	2,365	2,841	2,998	3,499	4,029	4,442	4,785	5,408
8	1,860	2,306	2,752	2,696	3,355	3,833	4,199	4,501	5,041
9	1,833	2,262	2,685	2,821	3,250	3,690	4,024	4,297	4,781
10	1,812	2,228	2,634	2,764	3,169	3,581	3,892	4,144	4,587
12	1,782	2,179	2,560	2,681	3,055	3,428	3,706	3,930	4,318
14	1,761	2,145	2,510	2,624	2,977	3,326	3,583	3,787	4,140
16	1,746	2,120	2,473	2,583	2,921	3,252	3,494	3,686	4,015
18	1,734	2,101	2,445	2,552	2,878	3,193	3,428	3,610	3,922
20	1,725	2,086	2,423	2,528	2,845	3,153	3,376	3,552	3,849
22	1,717	2,074	2,405	2,508	2,819	3,119	3,335	3,505	3,792
24	1,711	2,064	2,391	2,492	2,797	3,092	3,302	3,467	3,745
26	1,706	2,056	2,379	2,479	2,779	3,067	3,274	3,435	3,704
28	1,701	2,048	2,369	2,467	2,763	3,047	3,250	3,408	3,674
30	1,697	2,042	2,306	2,457	2,50	3,030	3,230	3,386	3,646
$\infty$	1,645	1,960	2,241	2,326	2,576	2,807	2,968	3,090	3,291

• слід урахувати наявність причинно-наслідкового зв'язку між показниками, що дає змогу розкрити сутність явищ, які

досліджуються;

- треба відбирати найбільш значущі фактори, оскільки охопити всі умови та обставини впливу на результативний показник неможливо;

- усі фактори мають бути кількісними з одиницями виміру;

- не рекомендується включати в кореляційну модель взаємопов'язані фактори, для чого їх слід перевірити на мультиколінеарність;

- не можна включати в кореляційну модель фактори, зв'язок яких з результативним показником має функціональний характер.

Розглянемо процедуру проведення багатфакторного кореляційного аналізу із застосуванням інструментів «Пакету аналізу» Excel.

**Етап 1.** Визначаємо фактори, що впливають на результативний показник, і відбираємо найістотніші з них. Основні правила відбору факторів:

- результативним фактором, як правило, визначається якісний показник ефективності певної сфери діяльності підприємства (прибуток, рентабельність, дохідність, обсяг реалізації, собівартість тощо);

- слід ураховувати наявність причинно-наслідкового зв'язку між показниками, що дає змогу розкрити сутність явищ, які досліджуються;

- треба відбирати найбільш значущі фактори, оскільки охопити всі умови та обставини впливу на результативний показник неможливо;

- усі фактори мають бути кількісними з одиницями виміру;

- не рекомендується включати в кореляційну модель взаємопов'язані фактори, для чого їх слід перевірити на мультиколінеарність;

- не можна включати в кореляційну модель фактори, зв'язок яких з результативним показником має функціональний характер.

Перевірка на мультиколінеарність передбачає оцінювання взаємозв'язку між окремими факторами-ознаками. За наявності лінійної залежності між факторами система нормальних рівнянь не матиме однозначного розв'язку, внаслідок чого коефіцієнти регресії та інші оцінки будуть нестійкими. Крім того, наявність взаємозв'язку факторів утруднює економічну інтерпретацію рівняння

зв'язку, оскільки зміна одного фактора спричиняє, як правило, зміну іншого, який з ним пов'язаний.

Існує кілька методів виключення мультиколінеарності, проте найчастіше застосовується метод оцінювання парних коефіцієнтів кореляції. Критерієм мультиколінеарності вважається виконання двох нерівностей:

$$\Gamma_{xjy} > \Gamma_{xjxk}; \Gamma_{xky} > \Gamma_{xjxk}$$

Якщо ці нерівності або хоча б одна з них не виконуються, то виключається той фактор  $x$ , зв'язок якого з результативним показником  $y$  буде менш тісним.

Проаналізуємо вплив різних факторів на валовий прибуток підприємства (табл. 2).

Для проведення аналізу було відібрано вісім факторів-ознак, кількість спостережень охоплювала 10 періодів.

Для оцінювання парного кореляційного зв'язку між факторами можна використати інструмент **Корреляция** з «Пакета аналізу» або статистичну функцію **KORREL**. У першому випадку дістанемо таблицю парних коефіцієнтів кореляції для кількох факторів одночасно (але без зворотного зв'язку з вхідними даними), у другому випадку можемо виконати обчислення лише для двох масивів.

Розглянемо спочатку процедуру застосування інструмента Корреляция. Показники, відібрані для проведення аналізу, оформляємо у вигляді таблиці-списку, дані якої можуть бути розміщені по рядкам або по стовпцям; першим рядком (стовпцем) списку є рядок Назв показників. Після подачі команди **Анализ данных.../Корреляция** з меню Сервис на екрані монітора з'явиться діалогове вікно, в якому потрібно вказати такі параметри:

- вхідний інтервал — посилання на діапазон аналізованих даних;
- групування — визначення послідовності розміщення даних (по рядках або по стовпцях);
- мітки — параметр для автоматичного формування рядка (стовпця) назв показників;
- вихідний діапазон — посилання на ліву верхню Клітинку вихідного діапазону активного робочого листка, нового робочого листка або нової робочої книги. При Цьому можна задати ім'я нового робочого листка, де вихідний діапазон почнеться з клітинки A1.

Проаналізуємо одержані результати, і відберемо фактори для подальшого аналізу:

- мають значний зв'язок з результативним фактором такі показники, як виробіток одного робітника, фондовіддача (зворотний зв'язок), матеріаловіддача (зворотний зв'язок) та витрати на 1 грн. реалізації (зворотний зв'язок), — перший стовпець табл. 3;

- є тісний зв'язок між фондо- і матеріаловіддачею, Виробітком та витратами на 1 грн. Виходячи з критерію Мультиколінеарності, слід відбракувати показник «фондовіддача». Показник «матеріаловіддача» також можна відбракувати, але його залишимо для ширшого аналізу;

- множина регресії включатиме аналіз впливу трьох факторів — виробітку одного робітника, матеріаловіддачі та витрат на 1 грн. реалізації.

**Зауваження.** Інструмент Корреляция є досить зручним у застосуванні, проте побудована таблиця коефіцієнтів не має зв'язку (посилань) на вхідну таблицю, тобто при зміні вхідних даних треба перераховувати коефіцієнти.

Одержати парні коефіцієнти кореляції можна також за допомогою статистичної функції **КОРРЕЛ**. Для цього поетапно розраховуємо кожну пару коефіцієнтів, при копіюванні формули фіксуємо посилання на адреси першого масиву відповідної пари (абсолютна адресація). В результаті дістанемо таку матрицю коефіцієнтів:

$YU$	$YX_1$	$YX_2$	...	$YX_m$
$UX_1$	$X_1X_1$	$X_1X_2$	...	$X_1X_m$
$UX_2$	$X_2X_1$	$X_2X_2$	...	$X_2X_m$
...	...	...	...	...
$UX_m$	$X_1X_m$	$X_2X_m$	...	$X_mX_m$

**Етап 2.** Будуємо рівняння множинної регресії та оцінюємо одержані результати. Для виконання цього етапу можна використати інструмент **Регресия** або статистичні функції.

Для роботи з інструментом **Регресия** вхідні дані Треба розмістити з дотриманням таких вимог:

- масиви даних розміщуються у стовпцях;

- перший рядок — назви показників;
- перший стовпець даних — масив  $y$  (результативний (фактор)), інші нерозірвані масиви  $x$  — масиви факторів-ознак.

Лінійний регресійний аналіз полягає у виборі графіка для відображення спостережень за допомогою методу найменших квадратів. Регресія використовується для аналізу впливу на залежну змінну значень однієї або більше незалежних змінних-факторів.

Параметрами діалогового вікна **Регресія** є:

- вхідний інтервал  $Y$  — посилання на діапазон результативного показника. Діапазон має складатися з одного стовпця;

- вхідний інтервал  $X$  — посилання на діапазон факторів-ознак. Максимальне число вхідних показників дорівнює 16;

- мітки — параметр для автоматичного формування назв показників;

- рівень надійності — дає змогу включити у вихідний діапазон рівень надійності до 95 %, що вводиться за замовчуванням;

- константа-нуль — прапорець, який вказує, що лінія регресії проходить через початок координат;

- вихідний діапазон — посилання на ліву верхню клітинку вихідного діапазону активного робочого листка, нового робочого листка або нової робочої книги. При цьому можна задати ім'я нового робочого листка, де вихідний діапазон почнеться з клітинки A1;

- залишки — дає змогу включити залишки у вихідний діапазон;

- стандартизовані залишки — дає можливість включити стандартизовані залишки у вихідний діапазон;

- графік залишків — діаграма залишків для кожної незалежної змінної;

- графік підбору — діаграма даних, що спостерігаються, а також прогнозованих значень для кожної незалежної змінної;

- графік нормальної імовірності — діаграма нормальної імовірності.

Вхідна таблиця для проведення регресійного аналізу валового прибутку включає чотири стовпці: результативний фактор і три фактори-ознаки, відібрані на першому етапі аналізу (табл. 4).

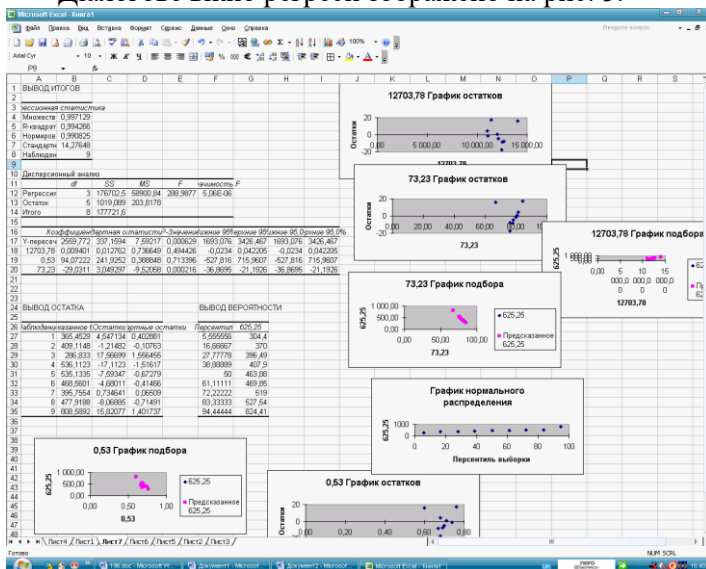


Як правило, виникає необхідність повернути діапазон вхідних даних, щоб вони були розміщені у стовпцях. Цю процедуру можна виконати кількома способами:

- без зв'язку з первинним діапазоном вхідних даних — скопіювати весь діапазон, вилучити зайві показники, виконати транспонування масиву (рядки у стовпці), подавши команду **Копировать, Специальная вставка/Транспонировать**;

- із зв'язком з первинним діапазоном вхідних даних — побудувати масив відібраних показників для аналізу за формулами-посиланнями на адреси первинного діапазону даних, вручну перемістити всі клітинки діапазону з відповідного рядка у стовпець за допомогою миші.

Діалогове вікно регресії зображено на рис. 5.



Таблиця Регрессионная статистика включає такі показники для оцінювання адекватності моделі:

- коефіцієнт детермінації;
- індекс кореляції;
- значення коефіцієнта детермінації при збільшенні кількості спостережень (нормоване);
- стандартну помилку;
- кількість спостережень.

Таблиця Дисперсионный анализ має таку структуру.

- $df$  — кількість ступенів вільності ( $m$ ;  $n-m-1$ ;  $n-1$ );
- $SS$  — дисперсія (факторна, залишкова, загальна);
- $MS$  — дисперсія/кількість ступенів вільності;
- $F$ - оцінка зв'язку між незалежними факторами і залежною змінною;
- значимість  $F$  — рівень значущості, що відповідає визначеному  $F$  — чим він нижче, тим кращий зв'язок.

Таблиця **Параметри моделі** має таку структуру:

- коэффициенты — значення параметрів моделі  $a_0, a_j$ ;
- стандартная ошибка — стандартна помилка параметрів рівняння;
- t-статистика — коефіцієнт/стандартна помилка,
- Р-значение — значущість для t-статистики;
- межі довірчих інтервалів для коефіцієнтів рівняння регресії при різних рівнях значущості.

Остання таблиця включає прогнозовані значення  $y$  і залишки.

При проведенні кореляційно-регресійного аналізу можна застосовувати також додаткові статистичні функції для оцінювання параметрів моделі та залежності між факторами:

- **НАКЛОН** — визначає коефіцієнт  $a_1$  у рівнянні  $y = a_0 + a_1x$ ,
  - **ОТРЕЗОК** — визначає коефіцієнт  $a_0$  у рівнянні  $y = a_0 + a_1x$ ,
  - **ЛИНЕЙН** — вводяться масиви  $y$  та  $x$  (можна декілька) — обчислюються коефіцієнти  $a_j$  і  $a_0$  або коефіцієнти у рівнянні з  $a_j$
- Вводити цю функцію у множинну регресію треба за правилами роботи з БД: виділити діапазон результатів (рядок із  $m + 1$  клітинок); ввести функцію; введення формули закінчити одночасним натисненням на клавіші  $\text{Ctrl}+\text{Shift}+\text{Enter}$ . Послідовність розміщення результатів відповідає параметрам моделі  $a_m, a_{m-1} \dots a_1, a_0$ ;

- **ПИРСОН** — визначає коефіцієнт кореляції  $R$  у межах від  $-1$  до  $+1$ ;
- **КВПИРСОН** — визначає коефіцієнт детермінації  $R^2$ ;
- **СТОШУХ** — визначає стандартну похибку прогнозних значень  $y$  для кожного значення  $x$  регресії;
- **КОВАР** — визначає коефіцієнти коваріації, а також середні попарні добутки відхилень.

### Хід роботи

Використовуючи інструмент Корреляция з «Пакета аналізу» Excel, оцінити кореляційну залежність (парну) від факторів (перший показник — результативний фактор, п'ять-шість показників за обґрунтованим вибором — незалежні змінні). Для розрахунку взяти дані з табл. 1 до теми 6. Результати (залишки, стандартизовані залишки та графіки) вивести на листку 7.

Із незалежних показників вибрати дві пари, що мають більш тісний кореляційний зв'язок з результативним фактором, та оцінити залежності на листку 8.

## **ПРАКТИЧНА РОБОТА №5**

### **ГРАФІЧНЕ ПОДАННЯ ДАНИХ**

**Мета:** ознайомитися із методами графічного подання статистичних даних

#### **Основні поняття**

Для зручності аналізу статистичних даних їх подають у вигляді графіка.

**Статистичний графік** — креслення, на якому за допомогою умовних геометричних фігур (ліній, точок або інших символічних знаків) зображено статистичні дані.

Правильно побудований графік подає аналітичну та статистичну інформацію більш виразно завдяки чому вона краще запам'ятовується. Графіки, даючи змогу виявити тенденції зміни показника, є обов'язковими при створенні презентацій, аналітичних документів, бізнес-планів тощо.

Можна виділити кілька видів графіків залежно від мети їх використання: порівняння, структурні, динамічні, зв'язку, контролю тощо.

Графічний метод подання та аналізу інформації є продовженням і доповненням табличного методу.

В Excel як загальну назву різних графіків, прийнято використовувати термін діаграма Excel. Кожна така діаграма складається з об'єктів, що виконують певні функції:

- вісь категорій — вісь, на якій показано заголовки категорій даних із таблиці;
- вісь значень — вісь, на якій розміщуються значення даних із таблиці;
- точка даних — елемент ряду даних, що відповідає значенню

однієї клітинки в таблиці;

- мітка даних — додається в діаграму для відображення конкретного значення точки даних;

- ряд даних — рядок або стовпець даних із таблиці;

- таблиця даних — таблиця, що розміщується нижче діаграми і містить вхідні дані для побудови діаграми;

- шкала — цифрові поділки на осі значень. Як правило, мінімальне значення відповідає числу нуль, а максимальне — найбільшому округленому з чисел, що зображуються;

- лінії сітки — лінії, що починаються з поділок шкали;

- область побудови діаграми — прямокутна область, обмежена осями;

- легенда — текстове поле з описом рядів даних і їх маркером;

- маркери виділення — з'являються на екрані монітора у вигляді чорних квадратиків при клацанні лівою клавішею миші по об'єкту під час вибору певного елемента діаграми;

- мітки поділок — маленькі лінії, що показують поділ шкали за категоріями та значеннями;

- заголовки — назва діаграми, заголовки осей значень і категорій;

- напис — текст, доданий до діаграми, який не залежить від даних;

- стрілка — графічний об'єкт, створений за допомогою панелі Рисование, який звертає увагу на певний елемент діаграми.

Використовують два види розміщення діаграм залежно від подальшого їх застосування:

- вміщені — діаграми поруч із табличними даними на одному робочому листку, що зручно, коли вони аналізуються разом;

- діаграмні листки — зручні при роздільному використанні їх і табличних даних. При цьому редактором автоматично формується стандартний листок заданого формату альбомної орієнтації.

Незалежно від того, яким способом створюватиметься діаграма, вона завжди динамічно пов'язана з даними листка й автоматично перебудовується при редагуванні даних.

**Редагування даних (значень)** можна здійснювати не тільки у вхідній таблиці, а й безпосередньо на діаграмах (двовимірних гістограмах, а також на лінійних, кругових, кільцевих, точкових та пазиркових діаграмах і графіках). Для цього потрібно виділити відповідну точку даних (клацнути по ній лівою клавішею миші —

виділиться ряд даних, а потім ще раз клацнути цією самою клавішею), підвести курсор миші до маркерів периметра, зловити момент перетворення маркера на подвійну стрілочку та збільшити або зменшити значення точки даних. При переміщенні редактор показуватиме одержане значення та його розміщення на осі У. Відповідного значення набуде точка даних у вхідній таблиці.

При використанні об'ємних діаграм іноді одні їхні елементи можуть перекривати інші. Щоб виправити таке становище, можна застосувати підняття діаграми і її поворот на деякий кут.

Для кращої наочності діаграми можна замінити зображення рядів даних або маркерів текстурою, штрихуванням, стандартними чи власними рисунками з цією метою треба:

- відкрити діалогове вікно Формат рядов данных, двічі клацнувши лівою клавішею миші по будь-якій точці даних вибраного ряду;
- перейти до розділу Вид й активізувати кнопку Способи заливки...;
- у новому діалоговому вікні вибрати відповідний розділ.

*Перший* його розділ Способи заливки дає змогу визначити кількість кольорів (один, два, стандартний різновид), вибрати конкретні кольори з палітри і тип штрихування (відносно розміщення кольорів — вертикальне, горизонтальне, перша діагональ, друга діагональ, із кутка, від центра).

*Другий* розділ діалогового вікна визначає вибір текстури заповнення ряду даних (папір, дерево, мармур, тканина тощо). Активізувавши кнопку Другая текстура... , можна відкрити діалогове вікно й вибрати відповідний файл з рисунком використовуваної текстури.

*Третій* розділ вікна визначає вибір типу і кольору штрихування (точки, горизонтальні або вертикальні лінії, а також лінії по діагоналі, подвійні лінії, штрихи хвилясті лінії, цегла тощо).

*Четвертий* розділ діалогового вікна дає змогу вибрати рисунок, розтягнути або розмножити діаграму. Наприклад, на діаграмі, зображеній на рис. 8.1, можна показати різні види виробів відповідними рисунками. Якщо вибрати опцію розмноження (чи розмноження в масштабі), то можна побачити кількість повних і неповних рисунків виробів.

Побудовану діаграму розмішують як на вільному місці листка, так і поверх даних. Переміщення та зміна розмірів діаграми здійснюються аналогічно стандартним процедурам роботи з графічними об'єктами.

Здебільшого при проведенні статистичного аналізу застосовуються динамічні ряди даних (досліджується зміна показників у часі). За необхідності можна побудувати графік залежності даних. Коротко розглянемо процедуру побудови графіка функції  $y(x)$ .

1. Створення таблиці вхідних даних. Для побудови графіка бажано мати певну кількість значень  $x$  (не менше 10). При цьому використовуються значення з однаковим кроком (створення арифметичної прогресії процедурою автозаповнення було розглянуто). Перше значення  $x$ , як правило, становить число нуль і графік починається від осі  $Y$  (команда Формат осі). У першій клітинці значень  $y$  записується відповідна формула або функція, яка копіюється з метою обчислення значень  $y(x)$  для всіх  $x$ .

2. Побудова діаграми типу графіка із згладженими лініями. На рис. 8.4 зображено побудовану діаграму функції  $y = \cos(x) + \sin(x)$ , де  $x[0;1]$ .

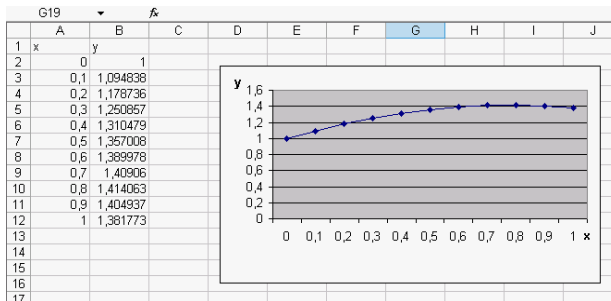


Рис. 4. Приклад графіка функції  $y(x)$

## ПРАКТИЧНА РОБОТА №6 ВИКОРИСТАННЯ МЕТОДІВ СТАТИСТИЧНОГО ОБРОБЛЕННЯ ДАНИХ

### Основні поняття

Масиви економічних показників характеризуються середніми значеннями величин різного виду, варіацією ряду, моментами і формою розподілу даних.

Для статистичного оцінювання даних використовують такі усереднені показники:

- середнє арифметичне;
- середнє квадратичне;
- середнє геометричне;
- середнє гармонічне;
- середнє кубічне.

Для оцінювання розсіювання (відхилення) даних відносно середнього значення користуються такими показниками:

- дисперсією;
- середньоквадратичним відхиленням — квадратним коренем із дисперсії (чим менше значення величини відносно її середнього значення, тим більш рівномірним є розподіл ряду даних);
- середнім модулем відхилень;

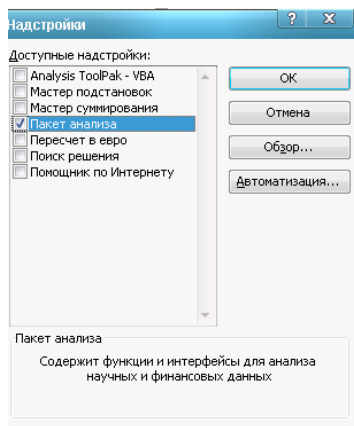
ексцесом (крутістю ряду або щільністю розподілу ймовірності).

Для нормального розподілу ексцес дорівнює нулю, для крутих кривих розподілу він додатний, для плоских — від'ємний порівняно з нормальною щільністю розподілу кривих;

• асиметричністю (ступенем асиметричності ряду або щільності розподілу ймовірності випадкової величини відносно її середнього значення). При симетричному розподілі коефіцієнт асиметрії дорівнює нулю;

- максимумом;
- мінімумом;
- найбільшим К-м (К — порядок значення, меншого за максимум);
- найменшим К-м (Е — порядок значення, більшого за мінімум);
- інтервалом (максимум — мінімум);
- модою (значенням, що найчастіше зустрічається у ряду даних);
- медіаною (значенням, розміщеним посередині ряду даних);
- квантилями розподілу (підмножинами даних з однаковим числом елементів);

Крім спеціальних статистичних функцій, для статистичного оцінювання даних можна використовувати спеціальний інструмент Описательная статистика з «Пакета аналізу» Excel.



Завантаження «Пакета аналізу» здійснюється за допомогою команди **Настройка...** з меню **Сервис**. У результаті на екрані монітора з'явиться діалогове вікно, в якому потрібно активізувати опцію **Пакет аналіза**.

«Пакет аналізу» включає великий набір інструментів для проведення статистичного, кореляційно-регресійного, дисперсного та інших видів аналізу. Подальший виклик інструментів «Пакета аналізу» можна здійснювати за командою **Анализ данных...** із меню **Сервис**.

Дані, за якими проводиться їх статистичний аналіз із використанням інструменту **Описательная статистика**, мають відповідати таким вимогам:

- 1) на листку вхідного діапазону даних не повинно бути об'єднаних клітинок;
- 2) один рядок (стовпець) назв показників має бути розміщений поруч із даними й утворювати з ними нерозірваний діапазон клітинок.

Наприклад, стовпець одиниці виміру можна розмістити перед стовпцем назв показників або наприкінці таблиці — крайній правий стовпець.



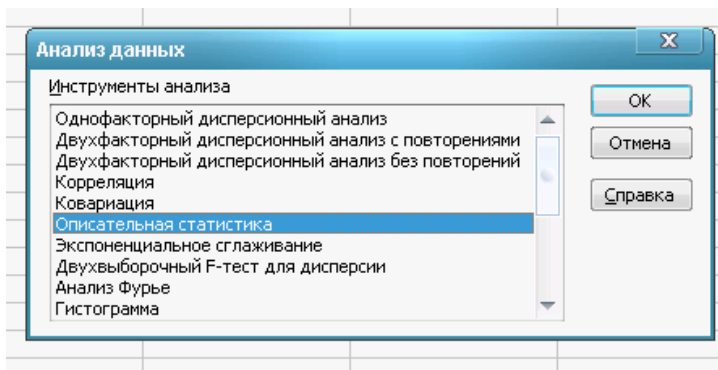


Рис. Вхідні дані для статистичного аналізу

Після подачі команди Анализ данных... Описательная статистика з меню Сервис на екрані монітора з'явиться діалогове вікно, в якому потрібно вказати такі параметри (рис. 2):

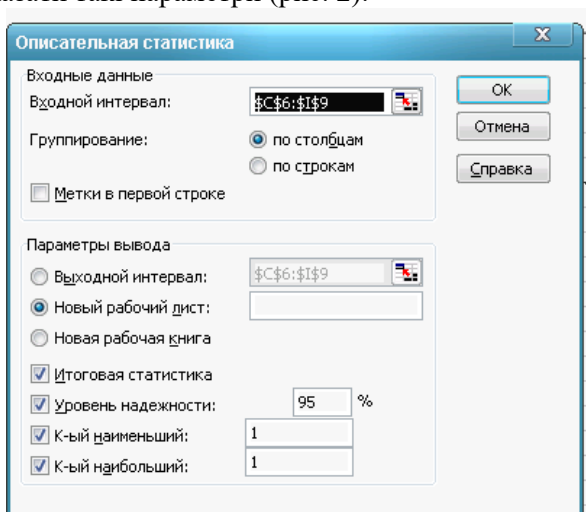


Рис. 2. Діалогове вікно Описательная статистика

- вхідний діапазон — посилання на клітинки, що містять аналізовані дані.

- групування — перемикач Группирование задає положення По столбцам або По строкам залежно від розташування даних у вхідному діапазоні;

- мітки в першому рядку (стовці) — перемикач задає положення назв показників Метки в первой строке, якщо перший рядок у

вхідному діапазоні назв містить Після цього можна продовжити оформлення таблиці, ввівши її назву, рамки, формати даних і вилучивши порожній рядок. У результаті дістанемо вихідну таблицю статистичних оцінок даних у вигляді, показаному нарис. 3.

Стовпці, або Метки в первом столбце, якщо назви рядків знаходяться у першому стовпці вхідного діапазону. Якщо вхідний діапазон не містить міток, то заголовки у вихідному діапазоні створюватимуться автоматично (Строка1\Столбец1);

- вихідний діапазон — посилання на ліву верхню клітинку вихідного діапазону;

- новий листок — перемикач активізують, щоб створити новий листок у книзі та помістити результати аналізу, починаючи з клітини A1. Якщо потрібно, то можна ввести ім'я нового листка в поле, розташоване навпроти відповідного положення перемикача;

- нову книгу — перемикач активізують, щоб створити нову книгу та помістити результати аналізу в клітинку A1 на першому листку в цій книзі;

- підсумкову статистику — прапорець активізують, якщо у вихідному діапазоні потрібно одержати додаткові статистичні оцінки даних: середнє, стандартну помилку (середнього), медіану, моду, стандартне відхилення, дисперсію вибірки, ексцес, асиметричність, інтервал, мінімум, максимум, суму;

- К-й найбільший — прапорець активізують, якщо у вихідну таблицю треба включити рядок К-го найбільшого значення для кожного діапазону даних;

- К-й найменший — прапорець активізують, якщо у вихідну таблицю потрібно включити рядок К-го найменшого значення для кожного діапазону даних;

- рівень надійності — прапорець активізують, якщо у вихідну таблицю треба включити рядок для рівня надійності. Наприклад, значенню 95 % відповідає рівень надійності середнього зі значущістю 0,05.

Розглянутий інструмент аналізу виводить два стовпці результатів для кожного показника даних. Лівий стовпець містить назви статистичних оцінок, а правий — статистичні оцінки. Відповідно над першим стовпцем розміщується назва показника, якщо було виділено рядок чи стовпець назв та активізовано перемикач Метки. Діапазон із двох стовпців буде виведений для кожного стовпця або

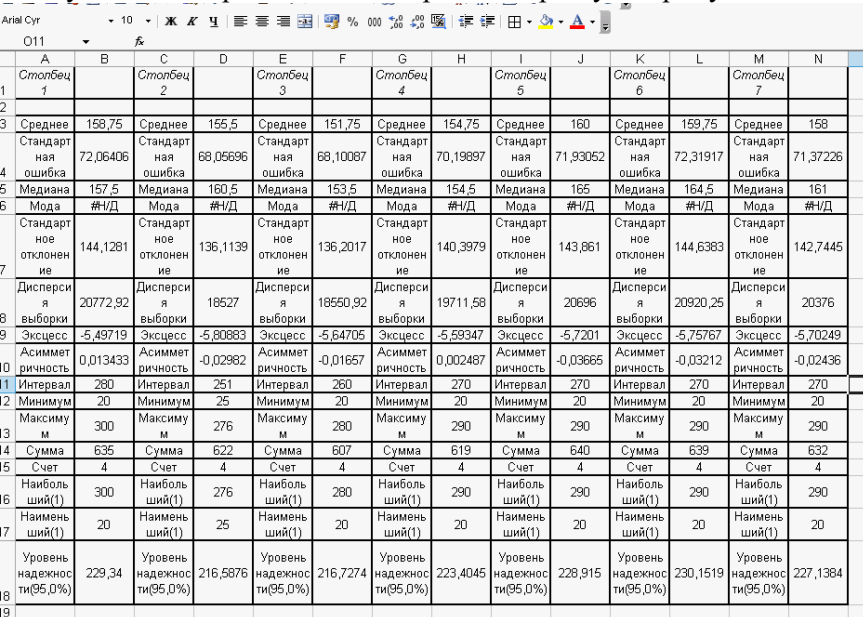
для кожного рядка вхідного діапазону показників залежно від положення перемикача Групування.

Для надання вихідній таблиці статистичного аналізу більш зручного вигляду треба послідовно виконати такі дії:

- перемістити назви показників у першому рядку із першого (третього, п'ятого і т. д.) стовпця у другий (четвертий, шостий і т. д.);

- вилучити дублюючі стовпці назв статистичних оцінок (третій, п'ятий і т. д.);

- розширити перший стовпець назв статистичних оцінок; задати для першого рядка назв показників формат розміщення тексту в кілька рядків та підібрати потрібну ширину стовпців.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Столбец 1		Столбец 2		Столбец 3		Столбец 4		Столбец 5		Столбец 6		Столбец 7	
3	Среднее	158,75	Среднее	155,5	Среднее	151,75	Среднее	154,75	Среднее	160	Среднее	159,75	Среднее	158
4	Стандартная ошибка	72,06405	Стандартная ошибка	68,05696	Стандартная ошибка	68,10087	Стандартная ошибка	70,19897	Стандартная ошибка	71,93052	Стандартная ошибка	72,31917	Стандартная ошибка	71,37226
5	Медиана	157,5	Медиана	160,5	Медиана	153,5	Медиана	154,5	Медиана	165	Медиана	164,5	Медиана	161
6	Мода	#И/Д	Мода	#И/Д	Мода	#И/Д	Мода	#И/Д	Мода	#И/Д	Мода	#И/Д	Мода	#И/Д
7	Стандартное отклонение	144,1261	Стандартное отклонение	136,1139	Стандартное отклонение	136,2017	Стандартное отклонение	140,3979	Стандартное отклонение	143,061	Стандартное отклонение	144,6383	Стандартное отклонение	142,7445
8	Дисперсия выборки	20772,92	Дисперсия выборки	18527	Дисперсия выборки	18550,92	Дисперсия выборки	19711,58	Дисперсия выборки	20696	Дисперсия выборки	20920,25	Дисперсия выборки	20376
9	Эксцесс	-5,49719	Эксцесс	-5,80883	Эксцесс	-5,64705	Эксцесс	-5,59347	Эксцесс	-5,7201	Эксцесс	-5,75767	Эксцесс	-5,70249
10	Асимметричность	0,013433	Асимметричность	-0,02982	Асимметричность	-0,01657	Асимметричность	0,002487	Асимметричность	-0,03665	Асимметричность	-0,03212	Асимметричность	-0,02436
11	Интервал	280	Интервал	251	Интервал	260	Интервал	270	Интервал	270	Интервал	270	Интервал	270
12	Минимум	20	Минимум	25	Минимум	20	Минимум	20	Минимум	20	Минимум	20	Минимум	20
13	Максимум	300	Максимум	276	Максимум	280	Максимум	290	Максимум	290	Максимум	290	Максимум	290
14	Сумма	635	Сумма	622	Сумма	607	Сумма	619	Сумма	640	Сумма	639	Сумма	632
15	Счет	4	Счет	4	Счет	4	Счет	4	Счет	4	Счет	4	Счет	4
16	Наибольший(1)	300	Наибольший(1)	276	Наибольший(1)	280	Наибольший(1)	290	Наибольший(1)	290	Наибольший(1)	290	Наибольший(1)	290
17	Наименьший(1)	20	Наименьший(1)	25	Наименьший(1)	20	Наименьший(1)	20	Наименьший(1)	20	Наименьший(1)	20	Наименьший(1)	20
18	Уровень надежности(95,0%)	229,34	Уровень надежности(95,0%)	216,5876	Уровень надежности(95,0%)	216,7274	Уровень надежности(95,0%)	223,4045	Уровень надежности(95,0%)	228,915	Уровень надежности(95,0%)	230,1519	Уровень надежности(95,0%)	227,1384

Рис. 5. Вихідна таблиця статистичних оцінок даних

Інструмент Описательная статистика досить зручний для здобуття 16 статистичних оцінок показників. Однак вихідна таблиця статистичного аналізу не є динамічною, оскільки включає лише результати обчислень (значення) і не зберігає зв'язку з таблицею вхідних даних. Якщо значення показників у вхідній таблиці

змінюватимуться, то кожний раз розрахунки треба виконувати заново.

Побудова таблиці статистичного аналізу за допомогою статистичних функцій займає набагато більше часу (кожну функцію необхідно вводити окремо). Проте вона дасть досить великий виграш у майбутньому при зміні вхідних даних.

Ще одним варіантом проведення статистичного аналізу може бути створення макроса для інструменту «Описова статистика».

### **СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ**

1. Інтегральні та комплексні оцінки стану навколишнього природного середовища : монографія / Васенко О. Г. та ін. Харків : НУГЗУ, 2015. 419 с.
2. Данилко В.К. Екологічна статистика: водні ресурси : монографія. Київ, 2003. 368 с.
3. Тарасова В.В. Екологічна статистика (з блочно-модульною формою контролю знань) : підручник. Київ : Центр учбової літератури, 2008. 392 с.
4. Данилко В. К. Екологічна статистика України: здобутки і проблеми. *Статистика України*. № 1, 2002. С. 54–57.
5. Веклич О. Підвищення стимулюючої ролі екологічного оподаткування в Україні. *Економіка України*. 2001. № 12. С. 29–37.